

# Weekly Report

## 01/19/2015 - 01/25/2015

Jing XIA

January 25, 2015

## 1 Summary

This week I mainly focus on the data inspection project.

## 2 Projects

### 2.1 Project 1 - Rank Visualization

-

### 2.2 Project 2 - Data Inspection

This week I implemented the algorithm of **highly correlated local region discovery** between two dimensions. Last week I proposed a simple method which expands entries with top 5% values to their neighborhoods. However, highly correlated local regions do not necessarily contains a large value, but might consist of a bunch of median values instead. The new algorithm I propose and implement is based on the graph-cut method, described as follows:

1. Initial: construct a graph which takes entries as nodes, neighborhood relations as edges and rmi (region mutual information) as edge weights.
2. Take every node  $v$  as a component, recording its own rmi.
3. Sort edges according to edge weight.
4. For each edge, if the two components of the two nodes which the edge connects are merged under two conditions: first the two nodes belong to different components, and second the rmi of the merging component is greater than a given threshold function  $f(x)$ .
5. Filter the component with rmi smaller than  $f(x)$

Besides, the threshold function is defined as

$$f(x) = \arctan(\frac{x-1}{3})/PI + min\_thresh$$

The resulting components covers all the regions with large rmi (proved). A similar method is widely adopted in image processing field for segmentation.

Next week I will start the quartet analysis implementation, which requires taking all views (one dimension views, two dimension views) together. The logic behind is a little complex though. The quartet analysis takes two optimization: embedding and partitioning to construct the C-tree.

Also, to speed up the whole process, we need to accelerate the computation with GPU. I will convert it during the winter vacation.

### **2.3 Project 3 - NBA Game Visualization**

-

## **3 Paper Reading**

-

## **4 Miscellaneous**

-

## **5 To Do List**

1. Data inspection project — quartet implementation.
2. Finish the top-k project next week.